

## Video 11 Fitting Lines to Data

1. How is the snowpack during wintertime in the Colorado Mountains measured? They have hundreds of sites in the Colorado mountains. They monitor it monthly. They do that by identifying sites that return to year after year. They collect a minimum of 10 depth and quantity subsections on the snowpack. Then they average those that come up with the total for the particular area.
2. What is a residual? A residual line minimizes the vertical distances of our data points from the line. As an example, using three points from the Colorado water data, we want to fit the best fitting-line with the most points as close as possible to it. The vertical distance of a point from a line is the residual. As we shift the line some residuals get smaller and some get smaller. They can never be all small at the same time. So the line needs to be where, as a group, they can be as small as possible
3. How does the least-squares method decide which line best fits the points in a scatterplot? Some of the residuals are positive and some are negative, the numbers are squared to have all positive numbers. If you add the squared residuals, the bigger the sum, the more the line misses the points. The sum needs to be as small as possible.
4. How can a particular year's data on the snowpack be used to predict the amount of water running downstream in the spring? Using  $\hat{y} = a + bx$ , the  $\hat{y}$  is the predicted value of  $y$ , not the measured value. We can use the regression line now to make predictions. The plotted points are our data, and the regression line predicts the  $y$ -value for any  $x$ . The regression line of the snowpack be useful to calculate the downstream of the spring. Using the least-squares method, you can subtract -7,920 from 1,941(30 inches of snowpack) to get the final calculation/prediction of knowing how much water is going to flow into the system in the spring. It gives us the forecast of 50,310 acre-ft of melted water.
5. The video showed two examples of residual plots. What does a residual plot tell you if the dots in the plot appear to be randomly scattered? What if the dots appear to form a strong curved pattern instead? The regression line works well to predict the water supply because it uses the relationship between snowpack and river flow and becomes linear. The dots that appear to be randomly scattered would be a good fit. You can use the residual using the least-square method. Using a scatterplot can help capture the data with clarity. But if the dots appear to be a strong curved pattern, it would not be a good fit since the best fit for the scatterplot would need to have a regression line. A curve won't make the greatest prediction for a regression line, and the linear model is the wrong one to use for a curved scatterplot.